

STAT107 Data Science Discovery

LAB: PLOTS

Man Fung (Heman) Leung

Fall, 2021

University of Illinois at Urbana-Champaign

- Please work in a group of 2–4 students
 - collaboration is important in data science!
 - meet new friends and discuss :)
 - let us know if you have any questions

Random fact of the day

The *Zen of Python* can be displayed by entering `import this`.

- Common/potential mistakes
 - “random_state” in `sample()`
 - “replace” in `sample()`
 - individual reflection(s) (2 pts each)
 - `concat()` vs `merge(..., how='outer')`
 - `concat()` before `sample()` in stratified random sampling
- Running the test cases successfully do not imply full score

- **Main page**
- Retrieve the lab using git
- Complete the notebook
 - pre-lab survey for 2 extra points
 - install matplotlib
 - hint for puzzle 2.2: use `["Salary"]["mean"].values[0]` after subsetting the correct department
 - hint for puzzle 3.1: try `groupby("Gender").plot.box()`
 - post-lab survey for 2 extra points
- Submit your work. Feel free to:
 - ask us questions
 - leave whenever you finish the lab

Default total number of cells: 61

- 0.1 in cell 7
- 0.a in cell 9 (reflection)
- 1.1 in cells 13–16
- 1.a in cell 19 (reflection)
- 1.2 in cell 21
- 1.3 in cell 23
- 2.1 in cell 26
- 2.2 in cell 29
- 2.3 in cell 32
- 2.a in cell 35 (reflection)
- 2.4 in cells 37–38
- 3.1 in cell 42
- 3.2 in cell 44
- 3.a in cell 48 (reflection)
- 4 in cells 50–51 (reflection)
- 5 in cell 55 (reflection)
- 6 in cell 58 (reflection)